Shared Bicycle Usage Analysis

Jairo Vera

2023-04-25

Introduction

This report provides a detailed analysis of the usage patterns of bike-sharing systems by user type (members and occasional users). Examine usage differences by day of week and time of day and train a Ridge regularized logistic regression model to predict user types based on travel duration and day of week.

Methodology

The following procedures were used for the analysis.

- Load and combine data files into a single dataset.
- Prepare and transform data to facilitate analysis.
- Generate usage plots by day of the week and time of day.
- Train and evaluate a logistic regression model with Ridge.
- Regularization to predict user type.

Results

Exploratory and Descriptive Analysis

The first step in the analysis is to look at the data to understand the differences in shared bike use by members and casual users.

Usage by user type and day of the week

The graph shows that the member uses the system more consistently throughout her week, while the occasional user uses it more frequently on weekends.

Usage by user type and time of day

The graph shows that member usage is more frequent during the morning and evening peaks, but occasional users peak in the afternoon and use the system more evenly throughout the day.

Predictive modeling

After examining the data, we use a variety of predictive modeling techniques to predict whether the ride was made by a member or the general public.

Data preparation

Before creating the model, we need to split the data into a training set and a test set and create a function to perform the regularization.

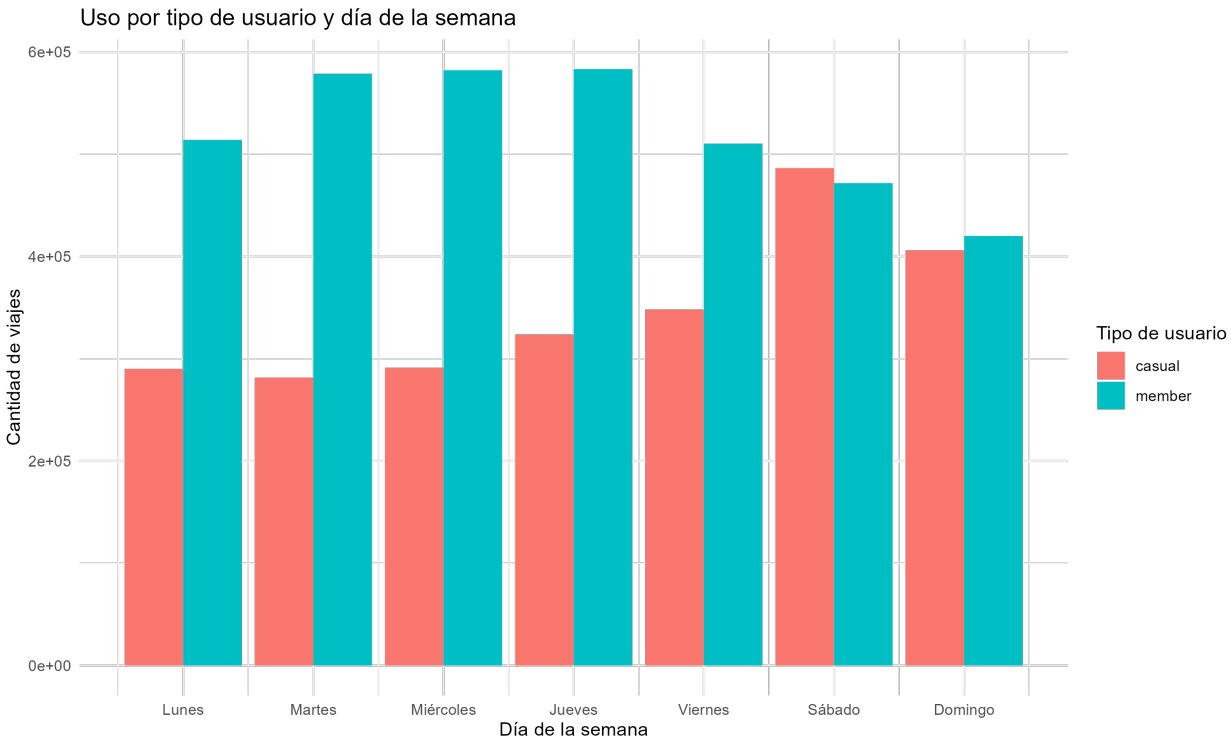

Figure 1: Usage by user type and day of the week

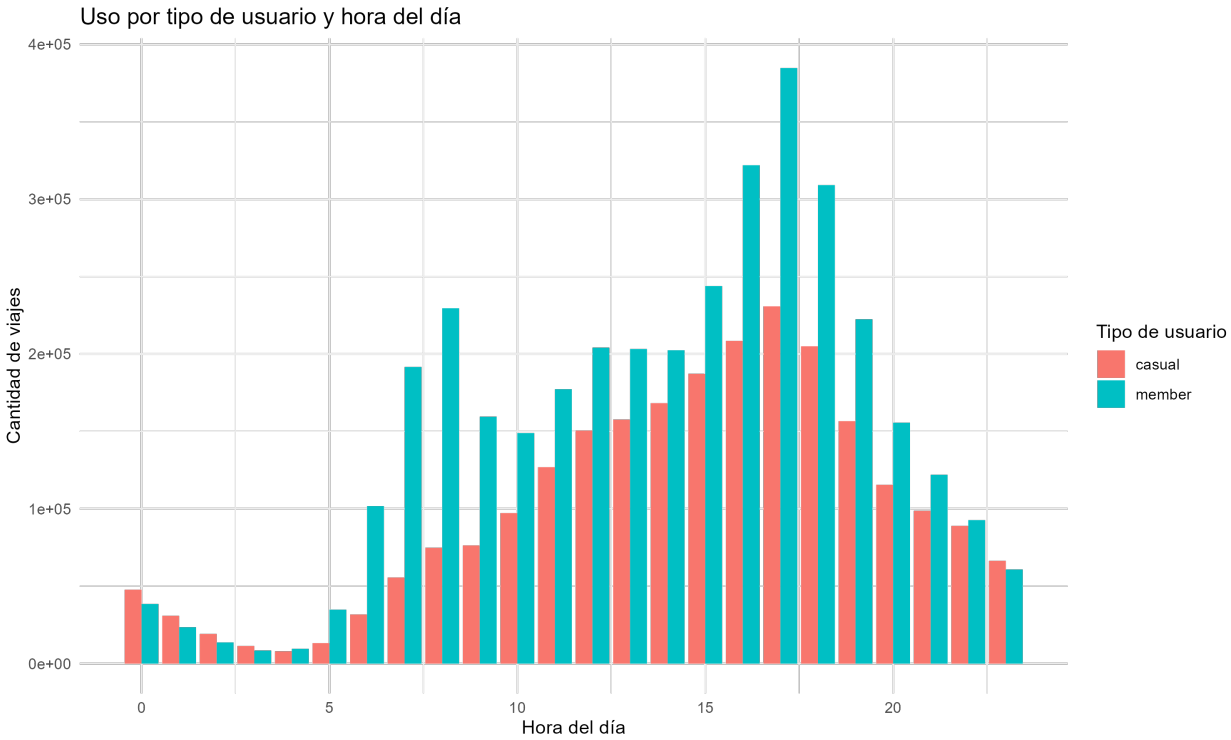

Figure 2: Usage by user type and time of day

Logistic Regression Model with Ridge Regularization

A logistic regression model with ridge regularization was trained with different lambda values. The best lambda value found is 0.001, giving the model an accuracy of about 63.74%. The confusion matrix and metrics of the final model show that the model has low sensitivity (15.64%) and high specificity (95.63%).

Discussion

The analysis revealed significant differences in the use of shared bike systems by user type, indicating opportunities for customized marketing and operational strategies. For example, we may offer targeted promotions to occasional users to encourage them to use the System during peak weekday hours when Members use the System more frequently.

The Ridge Regularized Logistic Regression model provides moderate accuracy in predicting user her type. However, the low sensitivity indicates that the model is not very effective at discriminating members. You can consider other modeling techniques and including more variables in your model to improve performance.

Conclusion

This analysis shows that there are notable differences in the use of shared bike systems between members and casual users. These results help inform marketing and operational strategies aimed at different user segments.

The Ridge Regularized Logistic Regression model has moderate accuracy but low sensitivity. Additional approaches may be required to improve the model's ability to accurately identify members. You can consider other modeling techniques and including more variables in your model to improve performance.

Overall, this analysis provides valuable insight into the behavior of bike-sharing system users and suggests ways to adapt the company's strategy to the needs of these different user segments. We recommend conducting further research and considering other modeling techniques to further improve our understanding of user behavior and the accuracy of user type predictions.